

# Will Epperson

I'm a Ph.D. student in the HCII at CMU advised by Dominik Moritz and Adam Perer.

I build interactive tools to help data scientists better understand and make decisions with their data by automating the tedious parts of analysis and letting analysts spend more time focused on data insights. Data quality issues are often "silent" – models will still train but predictions will be inaccurate or dashboards may unknowingly present inaccurate metrics, making data understanding and debugging a critical part of analysis. My research explores how to best support data debugging through tools that model user interest during analysis, augment their data programming environment with automatic visualization, and support reusing previous analysis workflows.

 willepperson.com

 willepp@cmu.edu

 CV PDF

 @w\_epperson

 @willeppy

 Google Scholar

 LinkedIn

---

## Education

August 2020 - Present

### Ph.D. in Human Computer Interaction

Carnegie Mellon University

*Advisors:* Dominik Moritz, Adam Perer

*Sample Coursework:* HCI Process and Theory, Computational Medicine, Human Judgement and Decision Making, Causality and ML, Advanced NLP

August 2020 - May 2023

### M.S. in Human Computer Interaction

Carnegie Mellon University

*Advisors:* Dominik Moritz, Adam Perer

August 2016 — May 2020

### B.S. in Computer Science

Georgia Institute of Technology

GPA: 4.0, Summa Cum Laude, threads in Intelligence and Modeling/Simulation

*Sample Coursework:* Machine Learning, Deep Learning, Computer Vision, Computer Architecture, Algorithms, Computer Simulation, Information Visualization

## Publications

### Dead or Alive: Continuous Data Profiling for Interactive Data Science

Will Epperson, Vaishnavi Gorantla, Dominik Moritz, Adam Perer

AutoProfiler is a Jupyter extension that helps data scientists understand their data and find issues during analysis through continuous data profiling.

*VIS 23: IEEE Conference on Data Visualization (VIS). Melbourne, Australia, 2023.*

 Project  PDF  Code  Best Paper Honorable Mention

## A Declarative Specification for Authoring Metrics Dashboards

Will Epperson, Kanit Wongsuphasawat, Allison Whilden, Fan Du, Justin Talbot

Quick dashboarding presents a novel specification for dashboard authoring, comprised of sections of metrics combined with dimensions.

*VDS at VIS 23: Visual Data Science Symposium (VDS). Melbourne, Australia, 2023.*

 Project  PDF  Best Paper

## Leveraging Analysis History for Improved In Situ Visualization Recommendation

Will Epperson, Doris Jung-Lin Lee, Leijie Wang, Kunal Agarwal, Aditya Parameswaran, Dominik Moritz, Adam Perer

Solas is a visualization recommendation tool that uses the history of analysis for in situ recommendations in Jupyter.

*EuroVis 22: Eurographics Conference on Visualization (EuroVis). Rome, Italy, 2022.*

 Project  PDF  Code  BibTeX

## Strategies for Reuse and Sharing among Data Scientists in Software Teams

Will Epperson, April Yi Wang, Robert DeLine, Steven M. Drucker

Interviews and a survey with 149 data scientists at Microsoft revealed five distinct strategies for sharing and reusing analysis code along with factors that encourage or discourage reuse.

*ICSE 22: ACM International Conference on Software Engineering (ICSE). Pittsburgh, PA, 2022.*

 Project  PDF  Recording  Slides  BibTeX

## Diff in the Loop: Supporting Data Comparison in Exploratory Data Analysis

April Yi Wang, Will Epperson, Robert DeLine, Steven M. Drucker

Diff in the Loop supports tracking, comparing, and visualizing differences in datasets during iterative data analysis.

*SIGCHI 22: ACM Symposium on Computer Human Interaction (CHI). New Orleans, LA, 2022.*

 Project  PDF  BibTeX

## RECAST: Interactive Auditing of Automatic Toxicity Detection Models

Austin P. Wright, Omar Shaikh, Haekyu Park, Will Epperson, Muhammed Ahmed, Stephane Pinel, Diyi Yang, Duen Horng (Polo) Chau

Interactive Auditing of Automatic Toxicity Detection Models

*24th ACM Conference on Computer-Supported Cooperative Work & Social Computing. 2021.*

 Project  PDF  BibTeX

## FairVis: Visual Analytics for Discovering Intersectional Bias in Machine Learning

Angel Cabrera, Will Epperson, Fred Hohman, Minsuk Kahng, Jamie Morgenstern, Duen Horng (Polo) Chau

Discovering intersectional ML Bias through interactive visualization.

*IEEE Conference on Visual Analytics Science and Technology (VAST). Vancouver, Canada, 2019.*

 Project  Demo  PDF  Blog  Recording  Code  BibTeX

# Talks

## Leveraging Analysis History for Improved In Situ Visualization Recommendation

June 2022

EuroVis 22: Eurographics Conference on Visualization

### **Strategies for Reuse and Sharing among Data Scientists in Software Teams**

May 2022

ICSE 22: ACM International Conference on Software Engineering

### **FairVis**

May 2019

VIS 19: IEEE Visualization Conference

## **Honors and Awards**

2019

PURA: President's Undergraduate Research Award

\$1500 research grant to continue work on FairVis project

2016

Stamps President's Scholarship

Full ride scholarship given to 40 incoming freshman at Georgia Tech

## **Research Experience**

August 2020 - Present

**Carnegie Mellon University**, Pittsburgh, PA

*Graduate Researcher, Data Interaction Group (DIG)*

Advisor: Dominik Moritz, Adam Perer

Member of the DIG research group, working on novel data visualizations, ML interpretation techniques, and interactive data systems.

*Relevant Skills:* Python, Javascript

January 2019 - May 2020

**Georgia Institute of Technology**, Atlanta, GA

*Undergraduate Researcher, Polo Club of Data Science*

Advisor: Duen Horng (Polo) Chau

Member of the Polo Club of Data Science working on novel data visualizations to find fairness issues in Machine Learning models

*Relevant Skills:* Python, Javascript

January 2018 - May 2019

**Georgia Institute of Technology**, Atlanta, GA

*Undergraduate Researcher, Automated Algorithm Design*

Advisor: Jason Zutty, Greg Rohling

Worked on EMAD algorithm design engine to implement sentiment analysis pipeline to analyze news articles to aid in predicting stock price movements using genetic algorithms. Led project to visualize the genetic algorithm evolution process.

*Relevant Skills:* Python, Javascript

## **Industry Experience**

Summer 2022

**Databricks**, San Francisco, CA

*Software Engineering Contractor*

Mentor: Kanit Wongsuphasawat

Designed and delivered production feature for creating dashboards by specifying fields of

interest in a dataset.

*Relevant Skills:* Typescript, Python

Summer 2021

**Microsoft Research**, Redmond, WA

*Research Intern, VIDA Group*

Mentor: Steve Drucker, Rob DeLine

Research intern working on data science tools. Lead project around reuse and sharing in data science, published at ICSE 2022. Also involved with project around visualizing data frame differences published at CHI 2022.

*Relevant Skills:* Python, Typescript

Summer 2019

**Point72 Asset Management**, New York, NY

*Data Analytics Intern, Market Intelligence Group*

Mentor: Trevor Rempel

Worked as Data Scientist in alternative data space to clean, model, and understand large datasets

*Relevant Skills:* Python, Distributed Computing in Spark

Summer 2018

**Ultimate Software**, Weston, FL

*Software Development Intern, Innovation Strategies Team*

Mentor: Joseph Cutrono

Designed and developed Slack app to integrate with the UltiPro HR management tool. App published to Slack app store.

*Relevant Skills:* Typescript, REST API development

Summer 2015

**The Home Depot**, Atlanta, GA

*Software Development Intern*

Developed web app for tracking candidate progress throughout hiring process for internal HR use.

*Relevant Skills:* Java, HTML/CSS/Javascript

## Mentees

During my PhD, I have had the pleasure of mentoring the following undergraduate students on research projects.

Summer 2021 - Fall 2021

**Leijie Wang**

Visualization recommendation for python in notebooks using history

Fall 2021 - Spring 2022

**Asad Sheikh**

Visualization recommendation for SQL using history

Spring 2022+

**Vaishnavi Gorantla**

Fact generation from data and presentation as text

## Teaching

Spring 2023

### **Graduate Teaching Assistant**

*Carnegie Mellon University, Pittsburgh, PA*

Programmable User Interfaces, Instructor: Scott Hudson

Taught recitation, designed assignments for class about UI design and intro HTML/CSS/Javascript.

Fall 2022

### **Graduate Teaching Assistant**

*Carnegie Mellon University, Pittsburgh, PA*

Interactive Data Science, Instructor: Adam Perer and John Stamper

Graded and office hours for class about using jupyter, visualization, steamlit and related tech for data science.

August 2017 - December 2018

### **Undergraduate Teaching Assistant**

*Georgia Institute of Technology, Atlanta, GA*

Intro to Database Systems (CS 4400), Instructor: Monica Sweat

Designed projects, held office hours and graded for relational databases class.

## **Service**

Reviewer for VIS 2023, CHI 2023, VIS 2022, CHI 2022, CSCW 2021, VIS 2021.

## **Leadership & Activities**

January 2019 - May 2020

### **Student Ambassador**

*Georgia Institute of Technology Alumni Association*

Serve as official representative of the Institute at events/tours for alumni, prospective students, and special guests.

January 2018 - May 2019

### **Executive Board Member -- Threads Co-chair**

*Stamps Scholars National Convention 2019*

Executive board member of Stamps Scholars National Convention, a 3-day conference with over 700 student attendees. Responsible for 20-person committee that plans and coordinates the different content threads of the convention.

## **Skills**

**Programing Languages:** Python (Advanced), Javascript/Typescript (Advanced), SQL (Intermediate), Java (Intermediate), C (Basic)

**Toolkits, Frameworks, Software:** Pytorch, Scikit-learn, Git, VegaLite, D3, Tableau, MacOS, Windows, Linux

**Natural Languages:** English (Native), Spanish (Advanced)